

Amplified restriction fragments for genomic enrichment (version 2.3 August 2011)

Tom Parchman (tparchma@uwyo.edu)
Zach Gompert (zgomper@uwyo.edu)
Alex Buerkle (buerkle@uwyo.edu)
University of Wyoming

This is a protocol for preparing highly-multiplexed samples for high-throughput sequencing. Genomic DNA is digested with restriction enzymes, barcodes and Illumina adaptor sequences are ligated to the digested fragments, and a subset of these fragments is amplified by PCR. Amplified fragments are purified and size-selected in agarose before being pooled into an Illumina sequencing library (see Fig. 2). This is a flexible protocol that has worked with a wide-variety of taxa and can be modified to target a variable number of fragments for sequencing (see Fig. 3).

Disclaimer: We cannot guarantee your success with this protocol. We also caution that this protocol and other approaches to multiplexed sequencing lead to immense amounts of data, with stochastic variation in sequence coverage across individuals and genetic regions. Importantly, we recommend making a comprehensive plan for analyzing the data before you follow this protocol. If you do not know what you will do with $> 10^8$ barcoded sequencing reads, then this protocol will not be useful.

1 Reagents and Equipment

1. EcoRI (NEB, 20,000 units/mL) R0101L
2. MseI (NEB, 10,000 units/mL) R0525L
3. T4 DNA ligase (NEB, 400,000 units/mL) M0202L
4. BioRad Iproof High Fidelity DNA polymerase Cat no. 172-5301 (\$424 for 250 μ L [500 units])
5. Molecular Biology grade agarose (suggestion: BioRad 161-3101)
6. 1 mg/mL BSA
7. 1 M NaCl
8. DMSO
9. QIAquick gel extraction kit Cat no. 28706(\$464 for 250)
10. Thermal Cyclers
11. 96 well plates and strip caps.
12. Agarose gel rig
13. Full plate centrifuge
14. UV light table and UV face shields
15. Nanodrop or similar spectrophotometer
16. other reagents listed below

2 Oligonucleotide Sequences

2.1 Adaptors

2.1.1 EcoRI

1. See Table S1 for barcoded EcoRI primer combinations. These adaptor sequences come in pairs and after annealing become one double-stranded adaptor. The barcodes residing in the adaptors were created using python scripts described in (Meyer & Kircher, 2010). To create variability in the initial bases for Illumina sequencing, we created barcodes that were 8, 9, and 10 bases in length. We constructed sets of barcodes so that all barcodes differ by a minimum of 4 bases. This facilitates correction of sequencing errors in barcodes. From the total set, we used an additional python script to select a subset that is balanced in the usage of nucleotides at each position. Further information on the python scripts can be found at (<http://bioinf.eva.mpg.de/multiplex/>).
2. Mix 1 μL of each oligo in a pair (100 μM stock) with 98 μL of water to make 100 μL of 1 pmole/ μL (1 μM) of annealed, double-stranded adaptor stock. Heat to 95°C for 5 minutes and slowly cool to room temperature. This annealing step only needs to be performed once for a stock solution of adaptor. Keep the set of adaptors organized in plate format that is convenient for later use in setting up reactions (e.g., Fig. 1).

2.1.2 MseI

Mix 10 μL of the MseI1 and MseI2 oligos (100 μM stock) with 80 μL of water to make 100 μL of 10 pmol/ μL (10 μM) stock. Heat to 95°C for 5 minutes and slowly cool to room temp to anneal oligos into double-stranded adaptor. This annealing step only needs to be performed once for a stock solution of adaptor.

These oligos are the same for each fragment. On an Illumina sequencer and with our configuration of the sequencing primer, sequencing only occurs from the EcoRI side, so there is no need to barcode the MseI oligos.

MseI1: 5' GCAGAAGACGGCATAACGAGCTCTTCCGATCTG 3'

MseI2: 5' TACAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG 3'

2.1.3 PCR primers

Mix 5 μL of the Illpcr1 and Illpcr2 oligos (100 μM stock) with 90 μL of water to make a working solution (2.5 μM of each oligo).

The asterisks between the first three bases in each PCR primer represent phosphothiolation (this is an option when ordering oligos). This prevents exonucleases from acting on the ends of the fragments, and thus protects the product prior to sequencing.

The last base of the Illpcr2 primer is a selective base and is arbitrarily **G**. Different and additional selective bases could be added, as desired. However, there is some evidence that the selective base does not function with high-fidelity polymerase and could be omitted.

Illpcr1 (Forward): 5' A*A*TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT 3'

Selective Illpcr2 (Reverse): 5' C*A*AGCAGAAGACGGCATAACGAGCTCTTCCGATCTGTAAG 3'

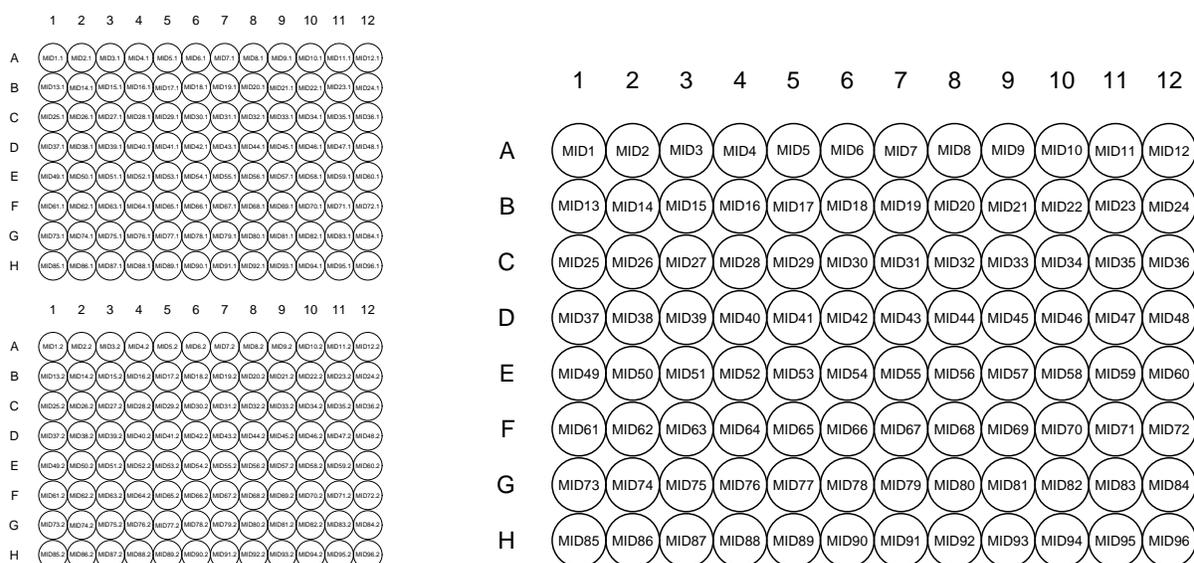


Figure 1: Sample layouts of 96 adaptor oligos on plates. At left are plates as they would be ordered. Oligos pairs are combined, annealed and then stored for use as in the plate on the right.

2.1.4 Illumina sequencing primer

This primer is used during Illumina sequencing, is not part of the library prep, and is presented here only to document its sequence.

5' ACACTCTTTCCCTACACGACGCTCTTCCGATC 3'

3 Restriction and Ligation Reactions

Genomic DNA is first digested with two restriction enzymes, EcoRI and MseI. This results in a pool of fragments that have the sticky-ended restriction cut sites on either end, and these ends provide a template for ligation of the customized adaptor sequences. The adaptor sequences contain the Illumina adaptors and primer sequences (and hence provide a binding site for the Illumina PCR primers). After the Illumina adaptor sequences on the EcoRI end, 8–10 base pair barcodes are incorporated that allow the identification of the origin of each sequencing read. The five bases at the 3' end of each adaptor oligo correspond to the ligation site; that is, they match the restriction cut site and have one added, protector base that prevents the cut from occurring again. The restriction digest is run first, followed by the permanent inactivation of enzymes by heating to 65°C. The ligation reaction is then run separately, so digestion of adaptors containing restriction enzyme cut sites cannot occur (a small number of our 768 EcoRI adaptor sequences contain the EcoRI or MseI cut sites; likewise for any other potential alternative restriction enzymes) and reactions can occur at their optimum temperatures.

3.1 Restriction Digest

1. Place 6 μL of sample DNA in each well of a plate (DNA should ideally be at a minimum concentration of 20 ng/ μL and a maximum concentration of 150 ng/ μL (note: this is for small

genomes, for large genomes add higher concentrations of DNA). Keep on ice.

2. For each sample prepare master mix I (Table 1), mix by vortexing, and centrifuge.

For these and all other reactions make sure to prepare an excess of mix to accommodate multiple rounds of pipetting, particularly if you are working with whole plates. Because the enzymes are stored in glycerol and other viscous solutions, a substantial volume is lost through adhesion to the outside of pipette tips. We suggest making 150% of what you think you will need.

3. Add 3 μL of the combined master mix I to each sample. Always keep cold once the enzymes have been added.
4. The total reaction volume should be 9 μL . Cover and seal the plate, vortex, centrifuge and incubate at 37 °C for 8 hours on a thermalcycler with a heated lid.

3.2 Adaptor Ligation

1. Thaw MseI and EcoRI adaptors. Have these adaptors annealed and easily accessible in plate format (for the EcoRI adaptors) (see Sections 2.1.1 and 2.1.2).
2. Add 1.4 μL of master mix II to each restriction-digested reaction.
3. Add 1 μL of EcoRI adaptors to each well on plate. Note that the MseI adaptor is in master mix I.
4. The total reaction volume should now be 11.4 μL . Cover and seal the plate, vortex, centrifuge and incubate at 16 °C for 6 hours on a thermalcycler.
5. Dilute the Restriction-Ligation reaction with 189 μL of 0.1 \times TE. Store at 4 °C for a month, or -20 °C for longer.

| Reagent | Number of samples 1 \times |
|-----------------------|---------------------------------|
| 10 \times T4 Buffer | 1.15 |
| 1M NaCl | 0.60 |
| 1 mg/mL BSA | 0.60 |
| Water | 0.25 |
| MseI | 0.12 |
| EcoRI | 0.28 |

Table 1: Reagents and volumes for Restriction Digest master mix I (3 μL prepared per sample).

4 PCR amplification

1. This PCR step uses the Illumina PCR primers to amplify fragments that have our adaptors+barcodes ligated onto the ends. To ameliorate stochastic differences in PCR production of fragments in reactions, we run two separate 20 μL reactions per restriction-ligation product, and later combine them.

| Reagent | Number of samples |
|---------------|-------------------|
| | 1× |
| MseI adaptor | 1 |
| Water | 0.072 |
| 10× T4 buffer | 0.1 |
| 1M NaCl | 0.05 |
| 1 mg/mL BSA | 0.05 |
| T4 DNA ligase | 0.1675 |

Table 2: Reagents and volumes for Restriction-Ligation master mix II (1.4 μ L prepared per sample).

| Reagent | 20 μ L Reaction | 40 μ L Reaction |
|----------------------------------|---------------------|---------------------|
| Water | 9.67 | 19.33 |
| 5× Iproof buffer | 4.0 | 7.99 |
| dNTP (10 mM) | 0.4 | 0.8 |
| MgCl ₂ (50mM) | 0.4 | 0.8 |
| Pre-mixed PCR Primers, 5 μ M | 1.33 | 2.66 |
| Iproof taq | 0.2 | 0.4 |
| DMSO | 0.15 | 0.30 |
| R/L product | 4 | 8 |

Table 3: Reagents and volumes for PCR amplification reactions.

2. Thermalcycler profile for this PCR: 98°C for 30s; 30 cycles of: 98°C for 20s, 60°C for 30s, 72°C for 40s; final extension at 72°C for 10 min.

5 Gel Purification

This step involves cutting a desired size range of fragments out of an agarose gel. The size range cut out of the gel can be adjusted to modify the number of fragments expected for sequencing. Because techniques such as this typically result in a negative relationship between fragment size and number, cutting out larger fragments should decrease the final number of fragments in the template and increase the coverage depth of these regions after sequencing. To eliminate variation in the size and amount of agarose punched out for each sample, it is recommended to pool all PCR reactions (they already have the barcodes incorporated) before loading the gel. The pooled library can then be run in as many wells on the gel as necessary. We are running 96 samples in replicate on two plates, pooling all of the PCR product, and then running approximately 10 lanes on a single agarose gel. If desired, allowing PCR product to evaporate on the bench top for several hours (or overnight) concentrates the PCR product, and means that fewer lanes are required on the gel to accommodate all of the samples (however, note that gels can be overloaded with concentrated DNA and lead to anomalous migration in the gel).

1. Run the PCR product out on 2.5% agarose gel at 100 volts for 2–3 hours (adjust voltage and time as needed for different gel boxes). It is important to include a good 100 bp or 50 bp ladder at multiple spots on the gel so that a clear line can be visualized across the

gel at a consistent size. Adding ethidium bromide into the gel (rather than staining after electrophoresis) reduces time and can improve consistency, and will not interfere with PCR after gel purification. We have been adding 9 μL of ethidium bromide per 200 mL of prepared gel. Be very careful that every gel is standardized as much as possible (same voltage, time, agarose concentration, etc.) in the case that multiple gels are prepped for a single library or project. This is necessary to ensure that the same size range is cut on every gel. Because PCR product can be entirely pooled for a given library, the most efficient way to proceed is to run 8–10 lanes of the pooled library on a thick gel, so 70–80 μL can be loaded into each well. This results in sufficient final library concentration and volume for Illumina sequencing (with ample excess).

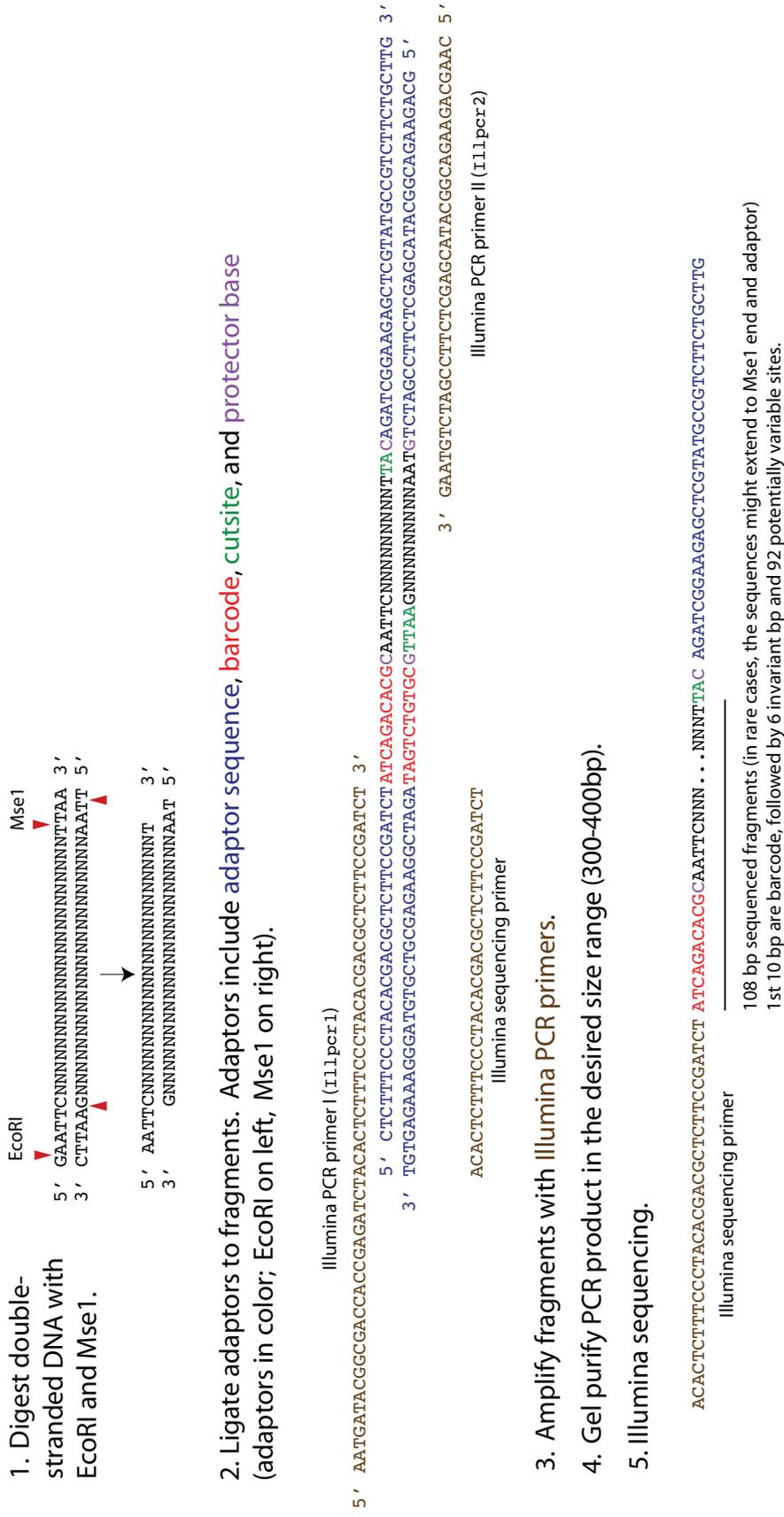
2. Cut the desired region out of the gel using the large end of sterile 1000 μL pipette tips. The region between 300–400 bp seems to be a good choice here and is recommended. Because of the potentially large amount of sequence representing Illumina PCR primers in the products, cut out fragments that are at least 250 bases in length. Because cutting out regions of a gel can take some time, be sure to use face and eye UV protection. In addition, realize that long periods of UV exposure to the DNA in the gel will have a mutagenizing effect. Therefore, minimize the amount of time that the gel is exposed to the UV.
3. Purify the excised gel punches using Qiaquick gel purification kits.

6 Preparing final template for Illumina sequencing

Use a Nanodrop or similar spectrophotometer to measure the DNA concentration of purified gel punch. Add the same amount of DNA from each gel purification to a 2 mL tube. A total concentration of 25 $\mu\text{L}/\text{ng}$ – 100 $\mu\text{L}/\text{ng}$ is ideal for sending off for Illumina sequencing, although library size requirements vary. We are taking 10–50 μL from each of the 10 purified gel punches and combining this into a pool to submit for sequencing. NCGR does not want to handle more than 500 μL , so submit samples in 500 μL tubes.

References

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**, pdb.prot5448–.



***Note only fragments with the EcoRI adaptor on one side and the MseI adaptor on the other will be sequenced due to the bridge-PCR method in Illumina sequencing.

Figure 2: A schematic of the steps involved in the Restriction-Ligation and PCR steps involved in creating the sequencing library.

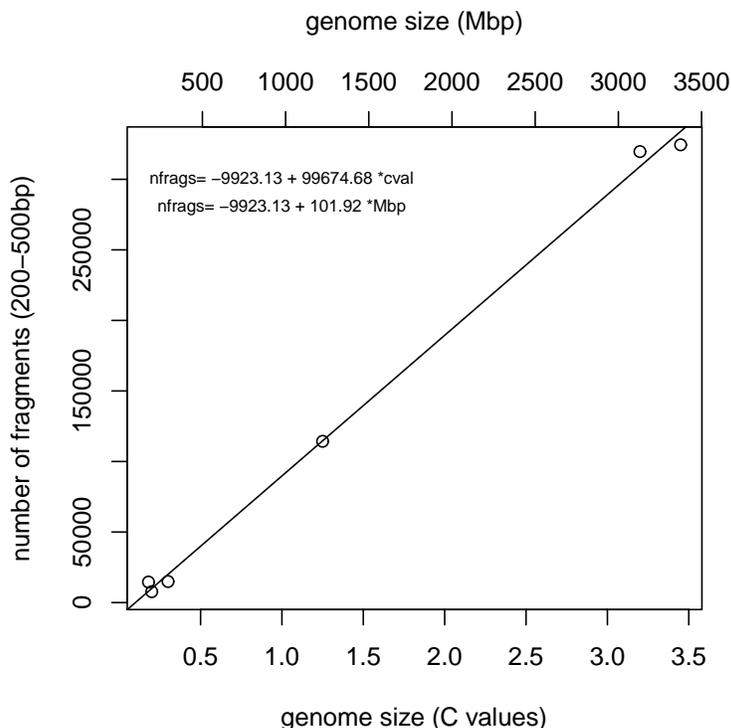


Figure 3: Number of fragments (200–500 bp) that are expected to be generated by digesting genomes of differing size. These expectations were derived by using Perl to digest reference genomes for *Drosophila*, chicken, mouse, *Arabidopsis*, *Tribolium* and human, and to count fragments within the 200–500 bp size range. In practice, a larger number of fragments is recovered and assembled in population studies.

Table S1: Example EcoRI adaptor sequences with barcodes added. The barcodes (in uppercase letters) are located between the adaptor sequence and the restriction site sequence (see Fig. 2). The full list of barcodes can be found in the supplemental file `annealed_barcode_out_768.csv`

| Name | Sequence (5' to 3') |
|--------|---|
| Mid1.1 | ctctttccctacacgacgctcttccgatctACGAGTGCCTc |
| MID1.2 | aattgACGCACTCGTtagatcggaagagcgtcgtgtagggaaagagtgt |
| MID2.1 | ctctttccctacacgacgctcttccgatctACGCTCGACAc |
| MID2.2 | aattgTGTCGAGCGTtagatcggaagagcgtcgtgtagggaaagagtgt |
| MID3.1 | ctctttccctacacgacgctcttccgatctAGACGCACTCc |
| MID3.2 | aattgGAGTGCCTCTtagatcggaagagcgtcgtgtagggaaagagtgt |
| MID4.1 | ctctttccctacacgacgctcttccgatctAGCACTGTAGc |
| MID4.2 | aattgCTACAGTGCTtagatcggaagagcgtcgtgtagggaaagagtgt |
| MID6.1 | ctctttccctacacgacgctcttccgatctATATCGCGAGc |
| MID6.2 | aattgCTCGGATATtagatcggaagagcgtcgtgtagggaaagagtgt |
| MID7.1 | ctctttccctacacgacgctcttccgatctCGTGTCTCTAc |

| Name | Sequence (5' to 3') |
|---------|---|
| MID7.2 | aattgTAGAGACACGagatcggagagcgtcgtgtagggaaagagtgt |
| MID8.1 | ctctttccctacacgacgctcttccgatctCTCGGTGTCc |
| MID8.2 | aattgGACACGCGAGagatcggagagcgtcgtgtagggaaagagtgt |
| MID10.1 | ctctttccctacacgacgctcttccgatctTCTCTATGCGc |
| MID10.2 | aattgCGCATAGAGAagatcggagagcgtcgtgtagggaaagagtgt |